GENOME ANNOUNCEMENT

Complete Genome Sequence of the Plant Pathogen *Erwinia amylovora* Strain ATCC 49946[▽]

M. Sebaihia,^{1†} A. M. Bocsanczy,^{2†} B. S. Biehl,³ M. A. Quail,¹ N. T. Perna,³ J. D. Glasner,³
G. A. DeClerck,⁴ S. Cartinhour,⁴ D. J. Schneider,⁴ S. D. Bentley,¹
J. Parkhill,¹ and S. V. Beer^{2*}

The Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom¹; Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, New York 14853²; Biotechnology Center, University of Wisconsin, Madison, Wisconsin 53706³; and United States Department of Agriculture, Agricultural Research Service, Tower Road, Ithaca, New York 14853⁴

Received 8 January 2010/Accepted 19 January 2010

***Erwinia amylovora* causes the economically important disease fire blight that affects rosaceous plants, especially pear and apple. Here we report the complete genome sequence and annotation of strain ATCC 49946. The analysis of the sequence and its comparison with sequenced genomes of closely related enterobacteria revealed signs of pathoadaptation to rosaceous hosts.**

Erwinia amylovora, a plant-associated member of the *Enterobacteriaceae*, causes fire blight, a devastating disease of rosaceous plants, especially pear and apple (6). The complete genome of Ea273 (ATCC 49946), a virulent strain isolated from an infected apple tree in New York State, was sequenced. Total DNA was extracted and prepared in pMAQ1 shotgun libraries. The complete shotgun sequence was obtained by using dye terminator chemistry in ABI 3730 automated sequencers and contains 88,457 reads (11.12-fold coverage), yielding a theoretical coverage of the genome of 99.99%. The sequence was assembled, finished, and annotated as described previously (1, 5), using Artemis (4) to collate data and facilitate annotation.

The genome of *E. amylovora* consists of a circular chromosome of 3,805,874 bp and two plasmids, AMYP1 (28,243 bp) and AMYP2 (71,487 bp). Coding regions in the chromosome account for 85.1% of the total sequence, with 3,483 identified coding sequences (CDS). Two hundred fifty-four (7%) of the CDSs do not have any matches in current NCBI databases; 114 (3.3%) correspond to conserved hypothetical proteins. Forty-nine CDSs (1.4%) are similar to genes from mobile elements such as integrases, transposases, and bacteriophages, and 110 CDSs (3.2%) were classified as pseudogenes due to interruptions or truncations of the CDSs. The remaining 2,956 annotated CDSs include among other categories genes involved in biosynthesis of the cellular envelope and modifications of surface proteins (299 CDSs [11%]) and genes involved in signal transduction and regulation (228 CDSs [8%]). Seven rRNA operons and 78 tRNA sequences were identified in the chro-

mosome; two new clusters were identified (AMY1550-1575 and AMY2648-2676) that resemble the T3SS-encoding SSR-1 island of *Sodalis glossinidius* (2), and four clusters that contain genes for biosynthesis of flagella, which based on their location might be regulated independently.

The smaller plasmid, AMYP1, had been reported as pEA29 (3); its sequence is nearly identical to the one reported here. The larger plasmid, AMYP2, renamed pEA72 for consistency in nomenclature, contains 87 predicted CDSs, with two predicted mobile-element-related CDSs and one pseudogene. Among the CDSs with annotated functions are a cluster of genes (AMYP2_49 to AMYP2_62) that encode a putative type IV fimbrial system (*pil* genes).

The genome of *E. amylovora* is only 3.8 Mb long, whereas most free-living enterobacteria, including plant pathogens, have genomes of 4.5 Mb to 5.5 Mb. Comparison of the genome of Ea273 with the sequenced genomes of 15 closely related enterobacteria identified 21 lineage-specific regions, which might be considered genomic islands. *E. amylovora* has many more predicted pseudogenes, relative to other enterobacteria with similar lifestyles. Given its size and the preponderance of pseudogenes, genome reduction may have occurred via mutational inactivation and subsequent deletion with the following consequences: *E. amylovora* has fewer genes involved in anaerobic respiration and fermentation than are found in typical related enterobacteria; this likely result in a reduced capacity to live in anaerobic environments.

The genome sequence of *E. amylovora* has revealed clear signs of pathoadaptation to the rosaceous plant environment. For example, T3SS-related proteins are present that are more similar to proteins of other plant pathogens than to proteins of closely related enterobacteria. These include type III effectors, homologous to those of plant-pathogenic pseudomonads, which confer virulence to *E. amylovora* in plants, and a sorbitol-metabolizing cluster that may confer a competitive advantage for survival in rosaceous plants. The reduced genome size

* Corresponding author. Mailing address: Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, New York 14853. Phone: (607) 255-7870. Fax: (607) 255-4471. E-mail: svbl@cornell.edu.

† M. Sebaihia and A. M. Bocsanczy are joint first authors.

▽ Published ahead of print on 29 January 2010.

and erosion or loss of genes involved in anaerobic respiration and nitrate assimilation are remarkable, relative to other plant- and animal-pathogenic members of the *Enterobacteriaceae*.

Nucleotide sequence accession numbers. The sequence and annotation of the *E. amylovora* chromosome and plasmids have been deposited in the EMBL database under accession no. FN666575, FN666576, and FN666577. The sequence and annotation also are available through the ASAP database (www.asap.ahabs.wisc.edu/asap/home.php).

This work was supported by NSF/USDA CSREES Microbial Genome Sequencing grant no. 2004-35600-14258.

REFERENCES

1. Bell, K. S., M. Sebaihia, L. Pritchard, M. T. G. Holden, L. J. Hyman, M. C. Holvea, N. R. Thomson, S. D. Bentley, L. J. C. Churcher, K. Mungall, R. Atkin, N. Bason, K. Brooks, T. Chillingworth, K. Clark, J. Doggett, A. Fraser, Z. Hance, H. Hauser, K. Jagels, S. Moule, H. Norbertczak, D. Ormond, C. Price, M. A. Quail, M. Sanders, D. Walker, S. Whitehead, G. P. C. Salmond, P. R. J. Birch, J. Parkhill, and I. K. Toth. 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl. Acad. Sci. U. S. A.* **101**:11105–11110.
2. Dale, C., T. Jones, and M. Pontes. 2005. Degenerative evolution and functional diversification of type-III secretion systems in the insect endosymbiont *Sodalis glossinidius*. *Mol. Biol. Evol.* **22**:758–766.
3. McGhee, G. C., and A. L. Jones. 2000. Complete nucleotide sequence of ubiquitous plasmid pEA29 from *Erwinia amylovora* strain Ea88: gene organization and intraspecies variation. *Appl. Environ. Microbiol.* **66**:4897–4907.
4. Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**:944–945.
5. Sebaihia, M., B. W. Wren, P. Mullany, N. F. Fairweather, N. Minton, R. Stabler, N. R. Thomson, A. P. Roberts, A. M. Cerdano-Tarraga, H. Wang, M. T. Holden, A. Wright, C. Churcher, M. A. Quail, S. Baker, N. Bason, K. Brooks, T. Chillingworth, A. Cronin, P. Davis, L. Dowd, A. Fraser, T. Feltwell, Z. Hance, S. Holroyd, K. Jagels, S. Moule, K. Mungall, C. Price, E. Rabbittowitsch, S. Sharp, M. Simmonds, K. Stevens, L. Unwin, S. Whithead, B. Dupuy, G. Dougan, B. Barrell, and J. Parkhill. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* **38**:779–786.
6. van der Zwet, T., and H. L. Keil. 1979. Fire blight. A bacterial disease of rosaceous plants, vol. 510. Science and Education Administration, USDA, Washington, DC.